PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G10L 5/06, 9/00	A1	(11) International Publication Number: WO 96/13830 (43) International Publication Date: 9 May 1996 (09.05.96)

(21) International Application Number: PCT/US95/13416

(22) International Filing Date: 19 October 1995 (19.10.95)

(30) Priority Data:
329,393 26 October 1994 (26.10.94) US

(71) Applicant: DICTAPHONE CORPORATION (U.S.) [US/US]; 3191 Broadbridge Avenue, Stratford, CT 06497-2559 (US).

(72) Inventor: SORENSEN, Jeffrey, S.; 59 Briarwood Drive, Seymour, CT 06483 (US).

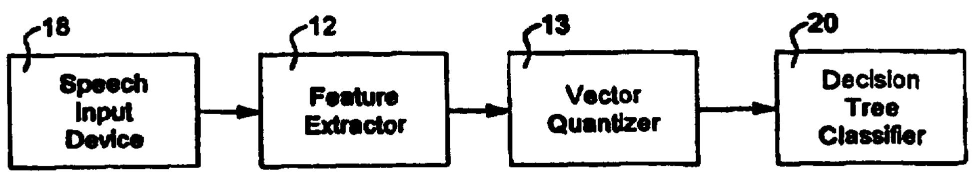
(74) Agents: NEFF, Gregor, N. et al.; Curtis, Morris & Safford, 530 Fifth Avenue, New York, NY 10036 (US).

(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TT, UA, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).

Published
With international search report.

(54) Title: DECISION TREE CLASSIFIER DESIGNED USING HIDDEN MARKOV MODELS

(57) Abstract



A decision tree classifier (20) is designed using the hidden Markov model to yield a final classifier that can be implemented without using any mathematical operations. The apparatus of this invention moves all of the mathematical calculations to the construction of the decision tree. Once this is completed, the decision tree can be implemented using only logical operations such as memory addressing and binary comparison operations. This simplicity allows the decision tree to be implemented in a simple hardware form using conventional gates.

DECISION TREE CLASSIFIER DESIGNED USING HIDDEN MARKOV MODELS

Field Of The Invention

This invention relates to pattern classification, and more particularly, to a method and apparatus for classifying unknown speech utterances into specific word categories.

Background Of The Invention

Speech recognition is the process of analyzing an acoustic speech signal to identify the linguistic message or utterance that was intended so that a machine can correctly respond to spoken commands. Fluent conversation with a machine is difficult because of the intrinsic variability and complexity of speech. Speech recognition difficulty is also a function of vocabulary size, confusability of words, signal bandwidth, noise, frequency distortions, the population of speakers that must be understood, and the form of speech to be processed.

A speech recognition device requires the transformation of the continuous signal into discrete representations which may be assigned proper meanings, and which, when comprehended, may be used to effect responsive behavior. The same word spoken by the same person on two successive occasions may have different characteristics. Pattern classifications have been developed to help in determining which word has been uttered.

Several methods of pattern classification have been utilized in the field of speech recognition. Currently, the hidden Markov model is the most popular statistical model for speech. The hidden Markov model characterizes speech signals as a stochastic state, or random distribution, machine with different characteristic distributions of speech signals associated with each state. Thus, a speech signal can be viewed as a series of acoustic sounds, where each sound has a particular pattern of harmonic features. However, there is not a one-to-one relationship between harmonic patterns and speech units. Rather, there is a random and statistical relationship between a particular speech

sound and a particular harmonic pattern. In addition, the duration of the sounds in speech does not unduly effect the recognition of whole words. The hidden Markov model captures both of these aspects of speech, where each state has a characteristic distribution of harmonic patterns, and the transitions from state to state describe the durational aspects of each speech sound.

The algorithms for designing hidden Markov models from a collection of sample utterances of words and sounds are widely known and disclosed in a book by Lawrence R. Rabiner and Biing Hwang Juang entitled "Fundamentals of Speech Recognition" Prentice-Hall, Inc. Englewood Cliffs, 1993, which is herein incorporated by reference. A method, commonly referred to as Baum-Welch re-estimation, allows one to continually refine models of spoken words. Once the statistical parameters of the model have been estimated, a simple formula exist for computing the probability of a given utterance being produced by the trained model. This latter algorithm is used in the design of isolated word speech recognition systems. By designing a collection of models, one for each word, one can then use the generated probability estimates as the basis for deciding the most likely word model to match to a given utterance.

The decision tree classifier is another pattern classification technique. Decision tree classifiers imply a sequential method of determining which class to assign to a particular observation. Decision tree classifiers are most commonly used in the field of artificial intelligence for medical diagnosis and in botany for designing taxonomic guides. A decision tree can be described as a guide to asking a series of questions, where the latter questions asked depend upon the answers to earlier questions. For example, in developing a guide to identifying bird species, a relevant early question is the color of the bird.

Decision tree design has proven to be a complex problem. The optimal decision tree classifier for a particular classification task can only be found by considering all possible relevant questions and all possible orders of asking them. This is a computationally impossible task even for situations where a small number of categories and features exist. Instead, methods of designing near optimal decision trees have been proposed using the measurements defined by the field of information theory.

Earlier attempts at designing decision tree classifiers for speech recognition using conventional methods proved unsuccessful. This is because an unrealistically vast quantity of training utterances would be required for accurate characterization of speech signals. This is due to the fact that the decision tree conditions questions based upon the responses to earlier questions. In any finite training set, and with each question, the number of examples of a particular utterance meeting all of the necessary criteria imposed by earlier questions dwindles rapidly. Thus, as observations accumulate with deeper nodes of the decision tree, the estimates used in designing the decision tree become increasingly inaccurate.

For the application of isolated word recognition, a specific class of hidden Markov models is employed. commonly referred to as discrete output hidden Markov models. For discrete output hidden Markov models, the spoken words are represented by a sequence of symbols from an appropriately defined alphabet of symbols. The method used to transform a set of acoustic patterns into a set of discrete symbols is known in the art as vector quantization.

Prior art implementations of speech recognition systems using hidden Markov models considered the contribution of all parts of the utterance simultaneously. The prior art required that the feature extraction and vector quantization steps be performed on the entire utterance before classification can begin. In addition, hidden Markov model computations require complex mathematical operations in order to estimate the probabilities of particular utterance patterns. These operations include the use of the complex representation of floating point numbers as well as large quantities of multiplications, divisions, and summations.

Summary Of The Invention

This invention overcomes the disadvantage of the prior art by providing a speech recognition system that does not use complex mathematical operations and does not take a great deal of time to classify utterances into word categories.

A decision tree classifier. designed using the hidden Markov model, yields a final classifier that can be implemented without using any mathematical operations. This invention moves all of the mathematical calculations to the construction, or design, of the decision tree. Once this is completed, the decision tree can be implemented using only logical operations such as memory addressing and binary comparison operations. This simplicity allows the decision tree to be implemented directly in a simple hardware form using conventional logic gates or in software as greatly reduced set of computer instructions. This differs greatly from previous algorithms which required that a means of mathematical calculation of probabilities be provided in the classification system.

This invention overcomes the problems of earlier decision tree classifier methods by introducing the hidden Markov modeling technique as an intermediate representation of the training data. The first step is obtaining a collection of examples of speech utterances. From these examples, hidden Markov models corresponding to each word are trained using the Baum-Welch method.

Once the models of the speech utterances have been trained, the training speech is discarded. Only the statistical models that correspond to each word are used in designing the decision tree. The statistical models are considered a smoothed version of the training data. This smoothed representation helps the decision tree design prevent over-specialization. The hidden Markov model imposes a rigid parametric structure to the probability space, preventing variability in the training data from ruining the tree design process. Estimates for unobserved utterances are inferred based on their similarity to other examples in the training data set.

In order to employ the information theoretic method of designing decision trees, it is necessary to derive new methods for estimating the probability values for specific sequences of acoustically labeled sounds. These new methods differ from earlier hidden Markov model algorithms in that they must yield probability values for sequences of sounds that are only partially specified. From a probabilistic perspective, the solution is to sum up the probabilities for all possible sequences of sounds for all sounds not

specified. However, this method is computationally prohibitive. Instead, a modification to the hidden Markov model algorithm known as the forward algorithm can be employed to compute the probabilities within a reasonable amount of time.

Using the techniques discussed above, it is possible to design the decision tree before the finical classification task. This is desirable, as this eliminates (virtually) all computations from the classification task.

Brief Description Of The Drawings

FIG. 1 is a block drawing of a speech recognition system that was utilized by the prior art;

FIG. 2 is a block diagram of the apparatus of this invention;

FIG. 3 is the decision tree;

FIG. 4 is a table of the time index and corresponding vector quantizer symbols for sample utterances; and

FIG. 5 is a block diagram of a decision tree design algorithm.

Description Of The Preferred Embodiment

Referring now to the drawings in detail and more particularly to FIG. 1: the reference character 11 represents a prior art speech recognition system for the words "yes", "no" and "help". The words "yes", "no" and "help" have been selected for illustrative purposes. Any words or any number of words may have been selected. Speech recognition system 11 comprises: a feature extractor 12; a vector quantizer 13; a hidden Markov Model (HMM) for the word "yes" 14; a hidden Markov Model (HMM) for the word "no" 15; a hidden Markov Model (HMM) for the word "help" 16 and a choose maximum 17.

The words "yes", "no" and "help" may be spoken into speech input device 18. Typically speech input device will consist of a microphone, amplifier and a A/D converter (not shown). The output of device 18 will be digitally sampled speech at a rate of approximately 12 Khz and a 16 bit resolution. The output of device 18 is coupled to the input of feature extractor 12.

Feature extractor 12 consists of a frame buffer and typical speech recognition preprocessing. Extractor 12 is disclosed in a book by Lawrence R. Rabiner and Biing Hwang Juang entitled "Fundamentals of Speech Recognition," Prentice-Hall, Inc., Englewood Cliffs, 1993, which is herein incorporated by reference. An implementation entails the use of a frame buffer of 45 ms with 30 ms of overlap between frames. Each frame is processed using a pre-emphasis filter, a Hamming windowing operation, an autocorrelation measurement with order 10, calculation of the Linear Prediction Coefficients (LPC's) followed by calculation of the LPC cepstral parameters. Cepstral parameters provide a complete source-filter characterization of the speech. The cepstral parameters are a representation of the spectral contents of the spoken utterance and contain information that includes the location and bandwidths of the formants of the vocal tract of the speaker. An energy term provides additional information about the signal amplitude. The output of extractor 12 are the aforementioned LPC cepstral parameters of order 10 and a single frame energy term that are coupled to the input of vector quantizer 13.

Vector quantizer 13 utilizes the algorithm disclosed by Rabiner and Juang in "Fundamentals of Speech Recognition" to map each collection of 11 features received from extractor 12 into a single integer. The above vector quantizer 13 is developed using a large quantity of training speech data. The integer sequence that constitutes a discrete version of the speech spectrum is then coupled to the inputs of HMM for the word "yes" 14, HMM for the word "no" 15 and HMM for the word "help" 16. HMM 14, HMM 15 and HMM 16 each individually contains a set of hidden Markov model parameters. When a sequence of integers representing a single spoken word is presented, the mathematical operations necessary to compute the observation sequence probability are performed. In the notation of speech recognition, this algorithm is known as the forward algorithm. HMM 14, HMM 15 and HMM 16 performs their own computations. These computations are time consuming.

The outputs of HMM 14, HMM 15 and HMM 16 are coupled to the input of choose maximum 17. Choose maximum 17 is a conventional recognizer that utilizes a computer program to determine whether HMM 14, HMM 15, or HMM 16 has the highest probability estimate. If HMM 14 had the highest

probability estimate, then the system concludes the word "yes" was spoken into device 18 and if HMM 15 had the highest probability estimate the word "no" was spoken into device 18. The output of choose maximum 17 is utilized to effect some command input to other systems to place functions in these other systems under voice control.

FIG. 2 is a block diagram of the apparatus of this invention. The output of speech input device 18 is coupled to the input of feature extractor 12 and the output of feature extractor 12 is coupled to the input of vector quantizer 13. The output of vector quantizer 13 is coupled to the input of decision tree classifier 20.

Decision tree classifier 20 has a buffer that contains the vector quantizer values of the entire spoken utterance, i.e. "yes", "no", "help". The decision tree classifier 20 may be represented by the data contained in the table shown in FIG. 3 and the associated procedure for accessing the aforementioned table. The table shown in FIG. 3 is generated using the statistical information collected in the hidden Markov models for each of the words in the vocabulary, i.e. "yes", "no" and "help". FIG 4 contains several examples of spoken utterances for illustrative purposes, and will be used in the discussions that follow.

The method used to classify utterances with decision tree classifier 20 involves a series of inspections of the symbols at specific time instances, as guided by the information in FIG 3. We begin with step S0 for each classification task. This table instructs us to inspect time index 3 and to proceed to other steps based on the symbols found at time index 3, or to announce a final classification when no further inspections are required. We will demonstrate this procedure for three specific utterances chosen from FIG 4.

We always begin the decision tree shown in FIG. 3 at step SO. This line of the decision tree tells us to look at the vector quantizer symbol contained at time index 3 (FIG. 4) for any spoken word. Let us assume that the word "no" was spoken. The symbol at time index 3 for the first word "no" is 7. The column for symbol 7 (FIG. 3) at time 3 tells us to proceed to step S3. This step tells us to look at time index 6 (FIG. 4). The word "no" has a symbol

7 at time index 6. The symbol 7 column of step S3 (FIG. 3) tells us to proceed to step S8. Step S8 tells us to refer to time index 9 (FIG. 4). At time index 9 for the word "no", we find a symbol value of 8, which informs us to go to column 8 of step 8 (FIG. 3). This tells us to classify the answer as a "no", which is correct.

If the word "yes" was chosen at step S0 of FIG. 3. This line of the table tells us to look at time index 3 (FIG. 4) for our chosen word. The symbol at time index 3 for the first sample utterance for the word "yes" is 7. The column for symbol 7 (FIG. 3) at time 3 tells us to proceed to step S3. This step tells us to look at time index 6 (FIG. 4). The word "yes" has a symbol 3 at time index 6. The symbol 3 column of step S3 (FIG. 3) tells us to classify this input, correctly, as the word "yes."

If the word "help" was spoken at step S0 of FIG. 3, this line of the table tells us to look at time index 3 (FIG. 4) for our chosen word. The symbol at time index 3 for the word "help" is 6. The column for symbol 6 (FIG. 3) at time 3 tells us to proceed to step S2. This step tells us to look at time index 4 (FIG. 4). The word "help" has a symbol 6 at time index 4. The symbol 6 column of step S2 (Fig. 3) tells us that the word "help" should be selected.

The "EOW" column of FIG. 3 tells us where to go when the end of a word has been reached at the time index requested for observation. The label "n/a" was included for utterances too short to qualify as any of the words in the training set.

The decision table shown in FIG. 3 can be used to correctly classify all of the examples presented in FIG. 4. However, the above decision tree and examples would not perform well on a large data set, as it is far too small to capture the diversity of potential utterances. Real implementations of the algorithm, which are hereinafter described, would generate tables with thousands of lines and at least several dozen columns. The above example is meant to illustrate the output of the hidden Markov model conversion to decision tree algorithm, and the procedure implemented in decision tree classifier 20 that uses the resultant data table.

FIG. 5 is a flow chart of a decision tree design algorithm. General decision tree design principles are disclosed by Rodney M. Goodman, and

Padhraic Smyth in their article entitled "Decision tree design from a communications theory standpoint", IEEE Transactions on Information Theory, Vol., 34, No. 5, pp. 979-997 Sept. 1988 which is herein incorporated by reference. The decision tree is designed using an iterative process. The precise rule applied for the expansion of each branch of the decision tree is specified. The desired termination condition depends highly on specific needs of the application, and will be discussed in broader terms.

In the preferred implementation, a greedy algorithm is specified where the termination condition is the total number of nodes in the resulting tree. This is directly related to the amount of memory available for the storage of the classifier data structures. Thus, this termination condition is well suited to most practical implementations.

In order to design the tree, the measure the mutual information between an arbitrary set of observations and the word category is needed. To find the mutual information, it is necessary to compute the probability of observing a partially specified sequence of vector quantizer labels. This is accomplished using a modified version of the hidden Markov model equations, contained in block 21 for the word "yes", block 22 for the word "no" and block 23 for the word "help", normally referred to as the forward algorithm. This modification involves removing the terms in the forward algorithm associated with unspecified vector quantizer outputs and replacing their probability values with one (1) in the conventional hidden Markov model algorithm.

This algorithm is contained in box 24 and is repeated iteratively, converting the terminal, or leaf, nodes of the decision tree into internal nodes one at a time, and adding new leaf nodes in proportion to the number of entries in the vector quantizer. This continues until the prespecified maximum number of tree nodes has been reached. The resulting data structures can then be organized into the, decision tree as has been shown in FIG. 3, and stored in the decision tree classifier 20.

The following decision tree design algorithm may be utilized in block 24: Before the decision tree algorithm can begin, a collection of discrete output hidden Markov models must be trained for each word in the recognition

vocabulary using a suitable training data set. The hidden Markov model parameters for a vocabulary of size W will be denoted as

$\mathbf{A}_i$ state transition matrix for word i

$\mathbf{B}_i$ state output distribution matrix for word i $\qquad i = \{1,2,\ldots,W$

π_i initial state distribution for word i

We further assume that the number of different outputs from the vector quantizer is denoted by Q, also known as the vector quantizer codebook size. In addition, we will define the notation for a set of vector quantizer outputs at different time indices as

$$X = \{(t_1,s_1),(t_2,s_2),\ldots,(t_n,s_n)\}$$

where the set X consists of pairs of observed symbols and their associated time index.

The tree design algorithm proceeds as follows: Two sets of sets are defined. The first set, T, is the collection of observation sets associated with internal tree nodes. The second set, L, is the collection of observation sets associated with leaf nodes of the tree. Initially, the set T is set to the empty set, and L is a set that contains one element, and that element is the empty set. The main loop of the algorithm moves elements from the set L to the set T one at a time until the cardinality, or number of elements, in set T reaches the pre-specified maximum number.

The determination of which set contained in L should be moved to set T at any iteration is determined using an information theoretic criterion. Here, the goal is to reduce the total entropy of the tree specified by the collection of leaf nodes in L. The optimal greedy selection for the collection of observation sets is given by

$$\max_{X \in L} \mathrm{H}(\underline{w}|X)\,\mathrm{P}(X)$$

where

with
$$\mathrm{H}(\underline{w}|X) = -\sum_{w} \mathrm{P}(w|X)\log_2 \mathrm{P}(w|X)$$

$$\mathrm{P}(w|X) = \frac{\mathrm{P}(X|w)\,\mathrm{P}(w)}{\mathrm{P}(X)} = \frac{\mathrm{P}(X|w)\,\mathrm{P}(w)}{\sum_{w} \mathrm{P}(X|w)\,\mathrm{P}(w)}$$

The probability of observing a collection of vector quantizer outputs conditioned on the word *w* is computed using the hidden Markov model equations. One expression of this calculation is given by

$$P(X|w=i) = \sum_{x_1, x_2, \ldots, x_{t_n}} \left(\prod_{j=1}^{n} \mathbf{B}_i[x(t_j), s_j] \right) \left(\pi_i[x(1)] \prod_{t=1}^{t_n - 1} \mathbf{A}_i[x(t), x(t+1)] \right)$$

which can be efficiently calculated using a slightly modified version of the hidden Markov model equation commonly referred to as the forward algorithm.

With each iteration, the optimal set *X* contained in *L* is moved to the set *T*. In addition, the optimal set *X* is used to compute the optimal time index for further node expansion. This is done using an information theoretic criteria specified by

$$\arg\min_{t \notin X} \sum_{s=1}^{Q} \mathrm{H}\big(w| X \cup \{(t,s)\}\big)\,\mathrm{P}\big(X \cup \{(t,s)\}\big)$$

The time index specified by this relationship is used to expand the collection of leaf nodes to include all possible leaf nodes associated with the set *X*. Each of these new sets, which are designated by

$$L \leftarrow L \cup \big\{X \cup \{(t,s)\}\big\} \qquad \forall s \in \{1,2,\ldots,Q\}$$

are added to the set *L* and the set *X* is removed from set *L*. Once this operation is completed, the algorithm repeats by choosing the best element of *L* to transfer to *T*, etc. This algorithm is expressed in pseudocode in FIG 6.

The above specification describes a new and improved apparatus and method for classifying utterances into word categories. It is realized that the above description may indicate to those skilled in the art additional ways in which the principles of this invention may be used without departing from the spirit. It is, therefore, intended that this invention be limited only by the scope of the appended claims.

WHAT IS CLAIMED IS:

1. A method of classifying utterance which comprises the steps of:

obtaining a collection of examples of speech utterances;

training hidden Markov models to obtain statistical models that correspond to individual words to classify the speech utterances;

using the statistical models to design a decision tree that represents a classifier of the speech utterances;

using the hidden Markov model to compute the probabilities of sounds not uttered; and

determining speech utterances with the decision tree.

2. The method claimed in 1, wherein the statistical models that are used to design the decision tree are implemented in a greedy iterative fashion.

3. The method claimed in Claim 2, wherein the recursive greedy fashion of implementing the decision tree design includes the steps of:

iteratively expanding entries of the decision tree table using the mutual information between the observed symbols and the word class; and

terminating the decision table when the maximum number of entries is reached.

4. The method claimed in Claim 3, wherein the mutual information is obtained in accordance with the equation

$$\begin{aligned} \mathrm{I}(\underline{w}; X) &= \mathrm{H}(\underline{w}) - \mathrm{H}(\underline{w} | X) \\ &= \mathrm{H}(\underline{w}) - \sum_{w} \mathrm{P}(w|X) \log_2 \mathrm{P}(w|X) \end{aligned}$$

5. The method claimed in Claim 3 further including the step of:

using the maximum expected information gain to determine the optimal time index to expand each entry in the decision tree table.

6. A system for classifying utterances which comprises:

means for imputing speech;

means coupled to said inputting means for transforming the speech into an alphabet of acoustic patterns;

means coupled to said transforming means for classifying the acoustic patterns into word categories by utilizing a decision tree classifier designed using hidden Markov models.

7. The system claimed in Claim 6, wherein said classifying means comprises:

a memory containing classification instructions;

means coupled to said memory for sequentially accessing said memory based upon the acoustic patterns; and

means coupled to said accessing means for outputting the terminal classification decision.

8. The system claimed in Claim 6, wherein said transforming means comprises:

a feature extractor for converting the inputted speech signals into frame based acoustical patterns; and

a vector quantizer coupled to said feature extractor for mapping the acoustical patterns to a fixed alphabet of symbols.

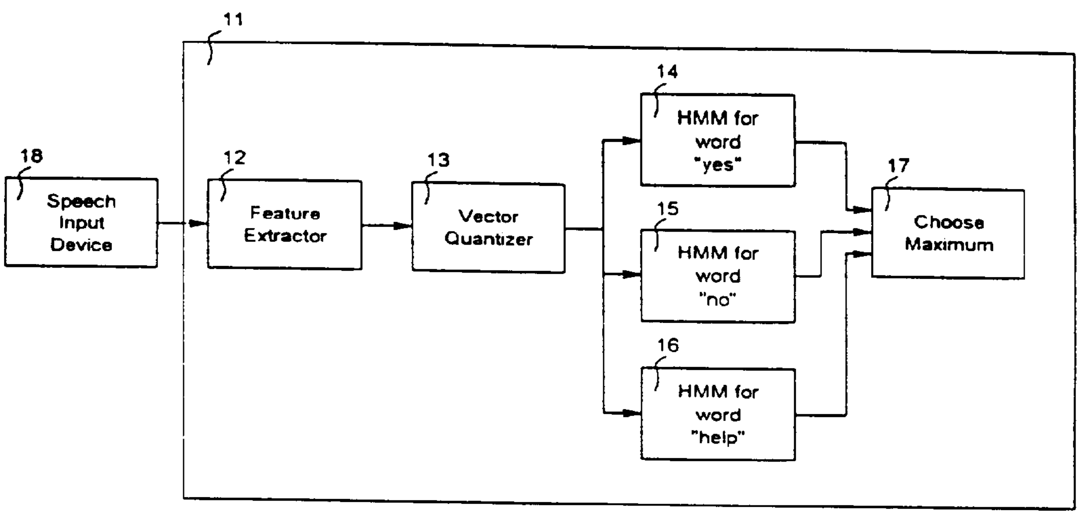


FIG. 1

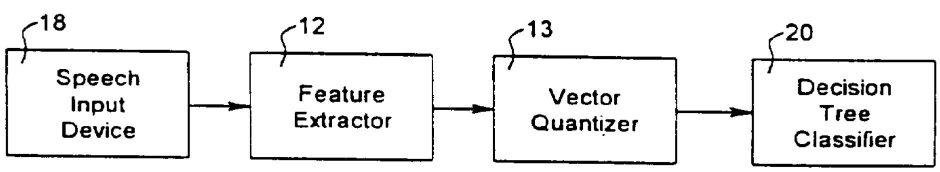


FIG. 2

		Vector Quantizer Symbol at Time Specified								
Step	Time	1	2	3	4	5	6	7	8	EOW
S0	3	no	yes	yes	S1	help	S2	S3	S4	yes
S1	1	help	help	no	S5	help	help	no	help	n/a
S2	4	help	help	yes	help	help	help	no	help	help
S3	6	no	S6	yes	no	no	S7	S8	no	no
S4	1	help	help	help	no	help	help	no	help	n/a
S5	2	no	no	yes	no	help	help	no	help	n/a
S6	1	yes	yes	yes	no	yes	yes	no	yes	n/a
S7	1	yes	yes	yes	no	yes	help	no	yes	n/a
S8	9	no	no	yes	no	no	no	no	no	no

FIG. 3

Sample Utterances	Time Index																																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
no	7	7	7	7	7	7	4	4	8	8	8	1	1	1	1	1	1	1	5	5	5	5	5	5	3	7												
no	7	7	7	7	7	7	7	4	4	8	8	8	1	1	1	1	5	5	5	5	5	5	5	3	7													
yes	7	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	7	2	2	2	2	2	2	2	2	2	2	2	2
yes	2	7	7	7	3	3	3	3	6	6	6	4	4	4	4	4	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2				
help	6	6	6	6	4	4	4	4	8	1	1	1	5	5	5	5	5																					
help	6	6	6	6	6	4	4	4	8	8	8	1	1	5	5	5	5																					

FIG. 4

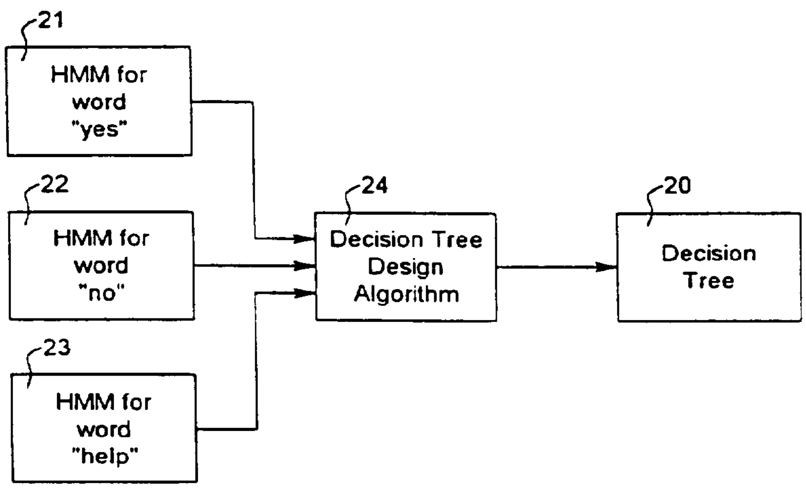


FIG. 5

```
DESIGNTREE( H, M )
    T ← {}
    L ←{ {} }
    while |T| < M
        do   X ← FINDMAXENTROPY( H, L )
             T ← T ∪ X
             L ← L - X
             t ← FINDBESTINFORMATIONGAIN( H, X )
             for s ← 1 to Q
                     do L ← L ∪ {(t,s)}
    return T
```

FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/13416

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :G10L 5/06, 9/00
US CL :395/2.65, 2.6
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 395/2.65, 2.6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y ----, P X	US,A, 5,455,889, 03 OCTOBER 1995, Figs. 1-7, cols. 1, 4, 5, 8.	1-5 ---- 6-8
Y	IEEE TRANSACTIONS ON INFORMATION THEORY, vol. 34, no.5, 01 SEPTEMBER 1988, Goodman et al., "DECISION TREE DESIGN FROM A COMMUNICATION THEORY STANDPOINT", pages 979-982.	1-5
Y	IEEE INTERNATIONAL WORKSHOP ON INTELLIGENT ROBOTS AND SYSTEMS, 05 NOVEMBER 1991, Moura-Pires et al., "DESIGN OF A DECSION TREE WITRH ACTION", pages 625-626.	4

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be part of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
04 JANUARY 1996	23 FEB 1996
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer JOHN MICHAEL GROVER
Facsimile No. (703) 305-3230	Telephone No. (703) 305-9600

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/13416

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, IEEE

search terms: decision tree, mutual information, classif?, greedy, forward, Baum, Welch, time index, Markov.